

AB-44

PCT

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/743,237

DATE: 08/31/2001

TIME: 12:30:19

Input Set : A:\84335127.app

Output Set: N:\CRF3\08312001\I743237.raw

3 <110> APPLICANT: SUGIHARA, TAKASHI  
 4 WADHWA, RENU  
 5 KAUL, SUNIL C.  
 6 MITSUI, YOUJI  
 8 <120> TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR  
 10 <130> FILE REFERENCE: 084335/0127  
 12 <140> CURRENT APPLICATION NUMBER: 09/743,237  
 13 <141> CURRENT FILING DATE: 2001-01-05  
 15 <150> PRIOR APPLICATION NUMBER: PCT/JP99/03859  
 16 <151> PRIOR FILING DATE: 1999-07-16  
 18 <160> NUMBER OF SEQ ID NOS: 24  
 20 <170> SOFTWARE: PatentIn Ver. 2.1  
 22 <210> SEQ ID NO: 1  
 23 <211> LENGTH: 2241  
 24 <212> TYPE: DNA  
 25 <213> ORGANISM: Mus musculus  
 27 <220> FEATURE:  
 28 <221> NAME/KEY: CDS  
 29 <222> LOCATION: (651)..(1535)  
 31 <400> SEQUENCE: 1

```

32 tatcctgtgg gttggccccg ggcagcaggc tctcagcagg ctcaggcacc acaagggata 60
34 cacagtgtgt gttcctggcc tgttgactt gtgactccac ccacctccgc cccagcaggg 120
36 ctagggatag aaccagggc cttttgcgtg ttctgcagat agtcttcagc ctggtagttt 180
38 ggggttggtt gggagatttt tttttcttca caccaaagac ttccattatt gaggattttt 240
40 tcagttgatg atctcccccc tctgtaagat aagggacagt tcttttaaacc tatgtagagt 300
42 tttgatgaat tctgcttctc aacctattg ctaagctata tagcaattcc ttgaaattgc 360
44 tatataactt aggagaacct ctgattctcc tgctctaca tcctgagtgc taggtgtaca 420
46 gggggaaatc attttggtga gactccgatg aactactgcc aggttcccaa ggcagcaagc 480
48 aagcaagaaa aagtgttgaa atcaaagaag caggtggtag tgtgccaggc ggcagccctg 540
50 aagacgcagc tttccaggcc cctctggctc aggaatcctg ttgcaagttc ccatcatccc 600
52 aggaggcaga ggaggcctcc agctgccctc ggaagaaaaga ctccagcccc atg gtg 656
53 Met Val
54 1
56 att tgt cag ctg aaa gga ggc gcc cag atg ctc tgc ata gac aac tgt 704
57 Ile Cys Gln Leu Lys Gly Gly Ala Gln Met Leu Cys Ile Asp Asn Cys
58 5 10 15
60 ggc gcg agg gag ctc aaa gcg ctc cat ctg ctt cct cag tac gat gac 752
61 Gly Ala Arg Glu Leu Lys Ala Leu His Leu Leu Pro Gln Tyr Asp Asp
62 20 25 30
64 cag agc agt ttc cct cag tca gag ctc cct aag cca atg aca act tta 800
65 Gln Ser Ser Phe Pro Gln Ser Glu Leu Pro Lys Pro Met Thr Thr Leu
66 35 40 45 50
68 gtg gga aga ctt ctg cca gta cca gcg aag tta aat ctc atc aca cag 848
69 Val Gly Arg Leu Leu Pro Val Pro Ala Lys Leu Asn Leu Ile Thr Gln
70 55 60 65
72 gtt gat aat gga gct ctc cca tca gct gtc aat ggg gct gcc ttt ccc 896
73 Val Asp Asn Gly Ala Leu Pro Ser Ala Val Asn Gly Ala Ala Phe Pro

```

ENTERED

## RAW SEQUENCE LISTING

DATE: 08/31/2001

PATENT APPLICATION: US/09/743,237

TIME: 12:30:19

Input Set : A:\84335127.app

Output Set: N:\CRF3\08312001\I743237.raw

```

74          70          75          80
76 tct gga cct gct ctg caa ggg cca ccc aaa ata act ctg tct ggg tac 944
77 Ser Gly Pro Ala Leu Gln Gly Pro Pro Lys Ile Thr Leu Ser Gly Tyr
78          85          90          95
80 tgt gac tgc ttc tcc agc ggg gac ttc tgc aac agc tgc agc tgc aac 992
81 Cys Asp Cys Phe Ser Ser Gly Asp Phe Cys Asn Ser Cys Ser Cys Asn
82          100          105          110
84 aac ctg cgc cat gag ctg gag cgc ttc aaa gcc ata aag gcg tgt ctt 1040
85 Asn Leu Arg His Glu Leu Glu Arg Phe Lys Ala Ile Lys Ala Cys Leu
86 115          120          125          130
88 gat aga aat cct gaa gct ttc caa cca aaa atg ggg aaa ggc cgt ctg 1088
89 Asp Arg Asn Pro Glu Ala Phe Gln Pro Lys Met Gly Lys Gly Arg Leu
90          135          140          145
92 gga gct gct aaa ctt cga cac agc aaa ggg tgc aac tgt aag cgc tca 1136
93 Gly Ala Ala Lys Leu Arg His Ser Lys Gly Cys Asn Cys Lys Arg Ser
94          150          155          160
96 ggc tgc ctg aag aac tac tgt gag tgc tat gag gcc aaa atc atg tgt 1184
97 Gly Cys Leu Lys Asn Tyr Cys Glu Cys Tyr Glu Ala Lys Ile Met Cys
98          165          170          175
100 tct tcc att tgc aaa tgc att gct tgc aaa aac tat gaa gaa agt cca 1232
101 Ser Ser Ile Cys Lys Cys Ile Ala Cys Lys Asn Tyr Glu Glu Ser Pro
102          180          185          190
104 gaa cga aaa atg ctg atg agc aca ccc cac tac atg gag cct ggg gac 1280
105 Glu Arg Lys Met Leu Met Ser Thr Pro His Tyr Met Glu Pro Gly Asp
106 195          200          205          210
108 ttt gag agc agc cat tat ttg tcc cca gcc aag ttc tca gga cct cca 1328
109 Phe Glu Ser Ser His Tyr Leu Ser Pro Ala Lys Phe Ser Gly Pro Pro
110          215          220          225
112 aaa ctg aga aaa aat agg cag gcc ttc tcc tgt atc tcc tgg gaa gta 1376
113 Lys Leu Arg Lys Asn Arg Gln Ala Phe Ser Cys Ile Ser Trp Glu Val
114          230          235          240
116 gtg gag gcc aca tgt gcc tgc ctg ctg gcc cag ggt gag gaa gca gag 1424
117 Val Glu Ala Thr Cys Ala Cys Leu Leu Ala Gln Gly Glu Glu Ala Glu
118          245          250          255
120 cag gag cac tgt tcc cca agc ttg gct gag cag atg atc ctg gag gag 1472
121 Gln Glu His Cys Ser Pro Ser Leu Ala Glu Gln Met Ile Leu Glu Glu
122          260          265          270
124 ttt gga agg tgc ctg tcg cag att ctc cac atc gag ttc aag tcc aag 1520
125 Phe Gly Arg Cys Leu Ser Gln Ile Leu His Ile Glu Phe Lys Ser Lys
126 275          280          285          290
128 ggg ctg aaa att gag tagcgtgcaa gctggtaaaag gggaatgcct gtggcaagcc 1575
129 Gly Leu Lys Ile Glu
130          295
132 tcagccctgg gaatctgcac cgaggaagct ggtgcccagg gaggagcaga ggccgcgcat 1635
134 catggccagg tcagctgtga ggtctgagtg atctgcatgg tactggccag cctactcaag 1695
136 gtatcctaaa gtgcaagcag gcagagccac cctggggatg gacactggcc ctctgtccc 1755
138 tggggaggcc ctctggggac tccctgccct gcataaaaag aggggtgattt tctacttgtt 1815
140 gttatgtgtt tgctttcaaa ttgcttagta gtacctccat tcaagttatt atgagccagc 1875
142 ctcaagttag agagctaggc tcttcttcag gtggactctg cccaaatcac atacaagtca 1935

```

## RAW SEQUENCE LISTING

DATE: 08/31/2001

PATENT APPLICATION: US/09/743,237

TIME: 12:30:19

Input Set : A:\84335127.app

Output Set: N:\CRF3\08312001\I743237.raw

```

144 ggtggccatc aggggttttt ccaggccagg cctgtgacag gagatatggg agggggggtcg 1995
146 ggtagagct ggggtttgtt ggattttttg cgtttttttc ttctgtatt tctgcttgaa 2055
148 gtgagaaaac ttgtctcctg tccaaccttt tctccataat tactgctgca cggtcgcctg 2115
150 ctgaccagtc acagtgacct cagacaccag aaggtgaggt ggcttattat gccacactt 2175
152 tgtgttttgt tgtgagaata aacctttcca gactcccaa aaaaaaaaaa aaaaaaaaaa 2235
154 aaaaaa 2241
157 <210> SEQ ID NO: 2
158 <211> LENGTH: 1861
159 <212> TYPE: DNA
160 <213> ORGANISM: Mus musculus
162 <220> FEATURE:
163 <221> NAME/KEY: CDS
164 <222> LOCATION: (271)..(1155)
166 <400> SEQUENCE: 2
167 ccagacgacg ccgctccgcc tcccgcctac agcgtgcacg tgttatcgtc gttacttccc 60
169 ggtgctcgcg ggcccgcgct gttgpcgcta agcgcaggag tgcgcgtgat cccagttgaa 120
171 atcaaagaag caggtggttag tgtgccaggc ggcagccctg aagacgcagc tttccaggcc 180
173 cctctggctc aggaatcctg ttgcaagttc ccatcatccc aggaggcaga ggaggcctcc 240
175 agctgccctc ggaagaaaga ctccagcccc atg gtg att tgt cag ctg aaa gga 294
176 Met Val Ile Cys Gln Leu Lys Gly
177 1 5
179 ggc gcc cag atg ctc tgc ata gac aac tgt ggc gcg agg gag ctc aaa 342
180 Gly Ala Gln Met Leu Cys Ile Asp Asn Cys Gly Ala Arg Glu Leu Lys
181 10 15 20
183 gcg ctc cat ctg ctt cct cag tac gat gac cag agc agt ttc cct cag 390
184 Ala Leu His Leu Leu Pro Gln Tyr Asp Asp Gln Ser Ser Phe Pro Gln
185 25 30 35 40
187 tca gag ctc cct aag cca atg aca act tta gtg gga aga ctt ctg cca 438
188 Ser Glu Leu Pro Lys Pro Met Thr Thr Leu Val Gly Arg Leu Leu Pro
189 45 50 55
191 gta cca gcg aag tta aat ctc atc aca cag gtt gat aat gga gct ctc 486
192 Val Pro Ala Lys Leu Asn Leu Ile Thr Gln Val Asp Asn Gly Ala Leu
193 60 65 70
195 cca tca gct gtc aat ggg gct gcc ttt ccc tct gga cct gct ctg caa 534
196 Pro Ser Ala Val Asn Gly Ala Ala Phe Pro Ser Gly Pro Ala Leu Gln
197 75 80 85
199 ggg cca ccc aaa ata act ctg tct ggg tac tgt gac tgc ttc tcc agc 582
200 Gly Pro Pro Lys Ile Thr Leu Ser Gly Tyr Cys Asp Cys Phe Ser Ser
201 90 95 100
203 ggg gac ttc tgc aac agc tgc agc tgc aac aac ctg cgc cat gag ctc 630
204 Gly Asp Phe Cys Asn Ser Cys Ser Cys Asn Leu Arg His Glu Leu
205 105 110 115 120
207 gag cgc ttc aaa gcc ata aag gcg tgt ctt gat aga aat cct gaa gct 678
208 Glu Arg Phe Lys Ala Ile Lys Ala Cys Leu Asp Arg Asn Pro Glu Ala
209 125 130 135
211 ttc caa cca aaa atg ggg aaa ggc cgt ctg gga gct gct aaa ctt cga 726
212 Phe Gln Pro Lys Met Gly Lys Gly Arg Leu Gly Ala Ala Lys Leu Arg
213 140 145 150
215 cac agc aaa ggg tgc aac tgt aag cgc tca ggc tgc ctg aag aac tac 774

```

## RAW SEQUENCE LISTING

DATE: 08/31/2001

PATENT APPLICATION: US/09/743,237

TIME: 12:30:19

Input Set : A:\84335127.app

Output Set: N:\CRF3\08312001\I743237.raw

```

216 His Ser Lys Gly Cys Asn Cys Lys Arg Ser Gly Cys Leu Lys Asn Tyr
217          155          160          165
219 tgt gag tgc tat gag gcc aaa atc atg tgt tct tcc att tgc aaa tgc 822
220 Cys Glu Cys Tyr Glu Ala Lys Ile Met Cys Ser Ser Ile Cys Lys Cys
221          170          175          180
223 att gct tgc aaa aac tat gaa gaa agt cca gaa aaa atg ctg atg 870
224 Ile Ala Cys Lys Asn Tyr Glu Glu Ser Pro Glu Arg Lys Met Leu Met
225 185          190          195          200
227 agc aca ccc cac tac atg gag cct ggg gac ttt gag agc agc cat tat 918
228 Ser Thr Pro His Tyr Met Glu Pro Gly Asp Phe Glu Ser Ser His Tyr
229          205          210          215
231 ttg tcc cca gcc aag ttc tca gga cct cca aaa ctg aga aaa aat agg 966
232 Leu Ser Pro Ala Lys Phe Ser Gly Pro Pro Lys Leu Arg Lys Asn Arg
233          220          225          230
235 cag gcc ttc tcc tgt atc tcc tgg gaa gta gtg gag gcc aca tgt gcc 1014
236 Gln Ala Phe Ser Cys Ile Ser Trp Glu Val Val Glu Ala Thr Cys Ala
237          235          240          245
239 tgc ctg ctg gcc cag ggt gag gaa gca gag cag gag cac tgt tcc cca 1062
240 Cys Leu Leu Ala Gln Gly Glu Glu Ala Glu Gln Glu His Cys Ser Pro
241          250          255          260
243 agc ttg gct gag cag atg atc ctg gag gag ttt gga agg tgc ctg tcg 1110
244 Ser Leu Ala Glu Gln Met Ile Leu Glu Glu Phe Gly Arg Cys Leu Ser
245 265          270          275          280
247 cag att ctc cac atc gag ttc aag tcc aag ggg ctg aaa att gag 1155
248 Gln Ile Leu His Ile Glu Phe Lys Ser Lys Gly Leu Lys Ile Glu
249          285          290          295
251 tagcgtgcaa gctggttaaag gggaatgcct gtggcaagcc tcagccctgg gaatctgcac 1215
253 cgaggaagct ggtgcccagg gaggagcaga ggccgcgcat catggccagg tcagctgtga 1275
255 ggtctgagtg atctgcatgg tactggccag cctactcaag gtatcctaaa gtgcaagcag 1335
257 gcagagccac cctggggatg gacactggcc ctctgtccc tggggaggcc ctctggggac 1395
259 tccctgccct gcataaaaag aggggtgatt tctacttgtt gttatgtgtt tgctttcaaa 1455
261 ttgcttagta gtacctccat tcaagttatt atgagccagc ctcaagttag agagctaggc 1515
263 tcttcttcag gtggactctg cccaaatcac atacaagtca ggtggccatc aggggttttt 1575
265 ccaggccagg cctgtgacag gagatatggg aggggggtcg ggttagagct gggtttgttt 1635
267 ggattttttg cgtttttttc ttcctgtatt tctgcttgaa gtgagaaaac ttgtctcctg 1695
269 tccaaccttt tctccataat tactgctgca cggtcgccctg ctgaccagtc acagtgcacct 1755
271 cagacaccag aagggtgaggt ggcttattat gccacactt tgtgttttgt tgtgagaata 1815
273 aacctttcca gactcccaaa aaaaaaaaaa aaaaaaaaaa aaaaaa 1861
276 <210> SEQ ID NO: 3
277 <211> LENGTH: 2134
278 <212> TYPE: DNA
279 <213> ORGANISM: Homo sapiens
281 <220> FEATURE:
282 <221> NAME/KEY: CDS
283 <222> LOCATION: (407)..(1303)
285 <400> SEQUENCE: 3
286 aattcgggggt caaggcgaag ctgcgggggg gcgacagcga cggcgggggag ctccctcgggg 60
288 agtaccgccg gatcccagag ctgagcgcg tgaggacgt cgcgctcctg caggccccgc 120
290 agccgccccg ctgcaacgtg cacttcctgt cctcgctgct acccgcgcac cgagccccgc 180

```

## RAW SEQUENCE LISTING

DATE: 08/31/2001

PATENT APPLICATION: US/09/743,237

TIME: 12:30:19

Input Set : A:\84335127.app

Output Set: N:\CRF3\08312001\I743237.raw

```

292 ggggtgttttg cccctggggc gcctgggtcc tgcgaaggag cctcccaccc gggcgtccgc 240
294 atgatcccag ttgaaatcaa ggtaagcagg tgggtactact acaagtaata atccggaaga 300
296 agcaactttg cagaatcttc ttgctcagga atcctgttgc aagttcccat ggtcccagga 360
298 actagaggat gcctcctgct gttctcttaa gaaagattcc aaccga atg gtg ata 415
299                                     Met Val Ile
300                                     1
302 tgc caa ttg aaa ggg ggc aca caa atg cta tgt ata gac aat tct aga 463
303 Cys Gln Leu Lys Gly Gly Thr Gln Met Leu Cys Ile Asp Asn Ser Arg
304      5                                10                                15
306 aca aga gaa cta aaa gca ctc cat ttg gtt cct cag tat caa gat caa 511
307 Thr Arg Glu Leu Lys Ala Leu His Leu Val Pro Gln Tyr Gln Asp Gln
308  20                                25                                30                                35
310 aat aat tat cta cag tca gat gtc cct aaa cca atg act gct tta gta 559
311 Asn Asn Tyr Leu Gln Ser Asp Val Pro Lys Pro Met Thr Ala Leu Val
312                                     40                                45                                50
314 ggg aga ttt ttg cca gca tca aca aaa tta aat ctc att aca caa caa 607
315 Gly Arg Phe Leu Pro Ala Ser Thr Lys Leu Asn Leu Ile Thr Gln Gln
316                                     55                                60                                65
318 ctt gag gga gcc tta cca tcg gta gtc aac ggg tct gct ttc ccc tcg 655
319 Leu Glu Gly Ala Leu Pro Ser Val Val Asn Gly Ser Ala Phe Pro Ser
320      70                                75                                80
322 gga tca act ctt cca gga cca aaa ata act ttg gct ggg tac tgt 703
323 Gly Ser Thr Leu Pro Gly Pro Pro Lys Ile Thr Leu Ala Gly Tyr Cys
324      85                                90                                95
326 gac tgc ttt gcc agt ggg gac ttt tgc aac aac tgc aat tgt aat aat 751
327 Asp Cys Phe Ala Ser Gly Asp Phe Cys Asn Asn Cys Asn Cys Asn Asn
328 100                                105                                110                                115
330 tgt tgc aac aac ttg cat cat gat att gaa cgg ttt aaa gcc att aag 799
331 Cys Cys Asn Asn Leu His His Asp Ile Glu Arg Phe Lys Ala Ile Lys
332                                     120                                125                                130
334 gca tgt ctt ggt aga aat cca gaa gct ttc cag cca aaa att ggg aag 847
335 Ala Cys Leu Gly Arg Asn Pro Glu Ala Phe Gln Pro Lys Ile Gly Lys
336                                     135                                140                                145
338 ggc caa ttg ggc aat gtc aag ccc cag cac aac aaa ggg tgc aac tgc 895
339 Gly Gln Leu Gly Asn Val Lys Pro Gln His Asn Lys Gly Cys Asn Cys
340      150                                155                                160
342 agg agg tca ggc tgc ctg aag aat tac tgc gag tgc tat gag gcc caa 943
343 Arg Arg Ser Gly Cys Leu Lys Asn Tyr Cys Glu Cys Tyr Glu Ala Gln
344      165                                170                                175
346 att atg tgt tct tct att tgc aaa tgc att ggt tgc aaa aat tat gaa 991
347 Ile Met Cys Ser Ser Ile Cys Lys Cys Ile Gly Cys Lys Asn Tyr Glu
348 180                                185                                190                                195
350 gaa agc cca gaa cga aag aca cta atg agc atg cca aac tac atg cag 1039
351 Glu Ser Pro Glu Arg Lys Thr Leu Met Ser Met Pro Asn Tyr Met Gln
352      200                                205                                210
354 act gga ggt ttg gaa ggc agc cat tac ctg cca cca acg aaa ttt tca 1087
355 Thr Gly Gly Leu Glu Gly Ser His Tyr Leu Pro Pro Thr Lys Phe Ser
356      215                                220                                225
358 gga ctt cca aga ttc agt cac gat agg cgg cct tcc tca tgc atc tcc 1135

```

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/743,237

DATE: 08/31/2001

TIME: 12:30:20

Input Set : A:\84335127.app

Output Set: N:\CRF3\08312001\I743237.raw

L:768 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23

**STATISTICS SUMMARY**

PATENT APPLICATION: US/09/743,237

DATE: 08/31/2001

TIME: 12:30:20

Input Set : A:\84335127.app

Output Set: N:\CRF3\08312001\I743237.raw

Application Serial Number: US/09/743,237

Alpha or Numeric: Numeric

Application Class:

Application File Date: 01-05-2001

Art Unit: PCT

Software Application: PatentIn

Total Number of Sequences: 24

Total Nucleotides: 6628

Total Amino Acids: 922

Number of Errors: 0

Number of Warnings: 1

Number of Corrections: 0

**MESSAGE SUMMARY**

341 W: 1 ((46) "n" or "Xaa" used)